

THE PERSONAL GENOME

Everyone has one

ADELAIDE RHODES, PH.D.

IMPORTANCE OF PERSONAL GENOMES

PERSONAL GENOMICS FOR HEALTHCARE WILL EXCEED CLOUD DEMAND FOR CLINICAL AND ACADEMIC GENOMICS RESEARCH

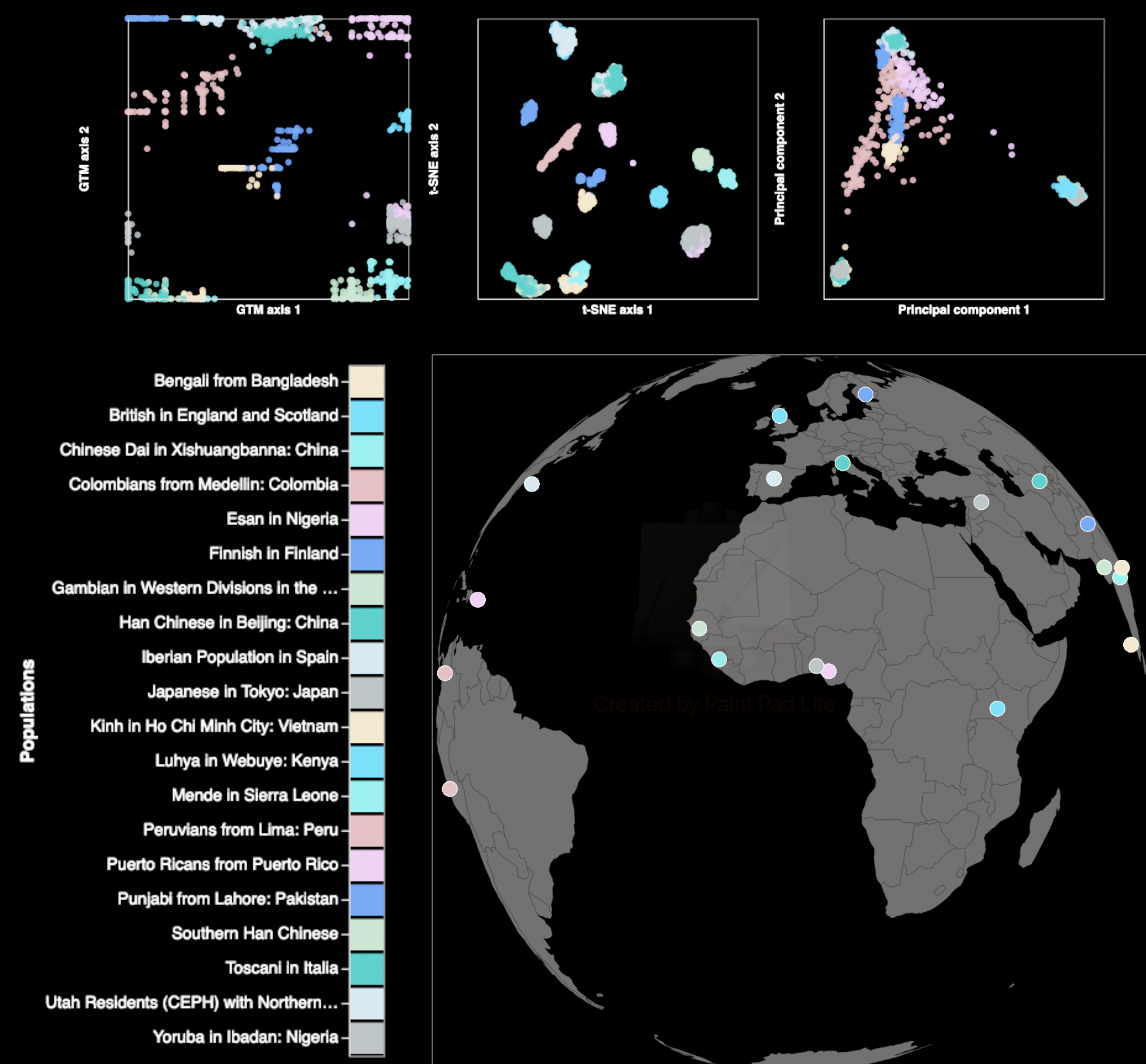
BY 2024
\$340 BILLION
DOLLARS/YEAR
WILL BE SPENT ON
CLOUD
COMPUTING

THE U.S. SPENDS
35%
OF OVERALL
WORLD FUNDING
ON GENOMICS
RESEARCH

BY 2025
60 MILLION
WILL HAVE THEIR
GENOME SEQUENCED
IN A HEALTHCARE
CONTEXT

Sources: <https://www.marketwatch.com/press-release/global-cloud-computing-market-size-2019-industry-trends-share-statistics-worldwide-overview-key-players-analysis-research-by-types-services-regional-outlook-and-forecasts-till-2024-2019-11-13>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2576262/>,
[https://www.cell.com/ajhg/fulltext/S0002-9297\(18\)30422-1](https://www.cell.com/ajhg/fulltext/S0002-9297(18)30422-1)

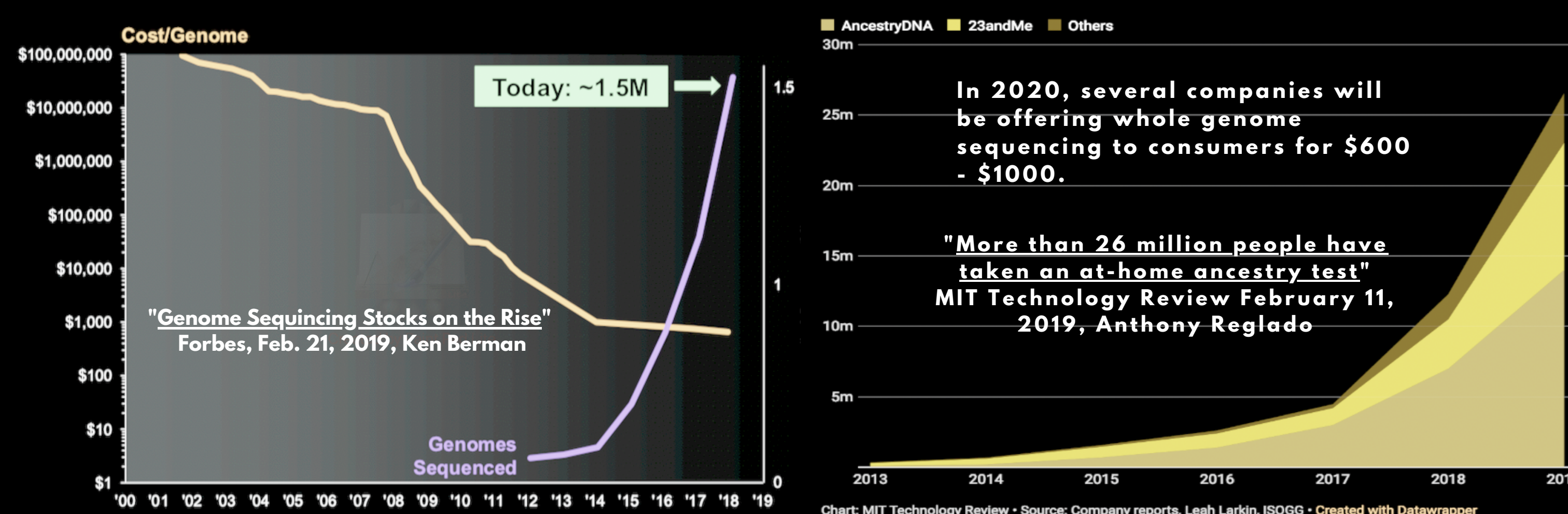
RESEARCHERS RELY ON INSIGHTS FROM PANGENOMES AND PERSONAL GENOMES



GASPAR, H.A., BREEN, G. PROBABILISTIC ANCESTRY MAPS: A METHOD TO ASSESS AND VISUALIZE POPULATION SUBSTRUCTURES IN GENETICS. BMC BIOINFORMATICS 20, 116 (2019). [HTTPS://DOI.ORG/10.1186/s12859-019-2680-1](https://doi.org/10.1186/s12859-019-2680-1)

BIOINFORMATIC CONSIDERATIONS

MASSIVE AMOUNTS OF DATA



- EACH GENOME GENERATES 100 GB DATA FOR DOWNSTREAM ANALYSIS, THIS REQUIRES STORAGE BEYOND WHAT THE TYPICAL HPC ACCOMMODATES
- 100 MILLION GENOMES X 100 GIGABYTES OF DATA = 10 EXABYTES OF DATA

NEED FOR INNOVATIVE TECHNOLOGY & CLOUD-NATIVE ANALYSIS

- NGS TOOLS ARE NOT CROSS-PLATFORM TOOLS. THE MAJORITY OF THEM WORK ON THE LINUX/UNIX OPERATING SYSTEM.
- MOST OF THE NGS TOOLS INTERACT WITH USERS VIA COMMAND LINE INTERFACE.
- VERY FEW NGS TOOLS HAVE BEEN DEVELOPED DURING THE LAST 5 YEARS.
- THERE IS NO STANDARD INPUT/OUTPUT FILE FORMAT FOR NGS TOOLS.

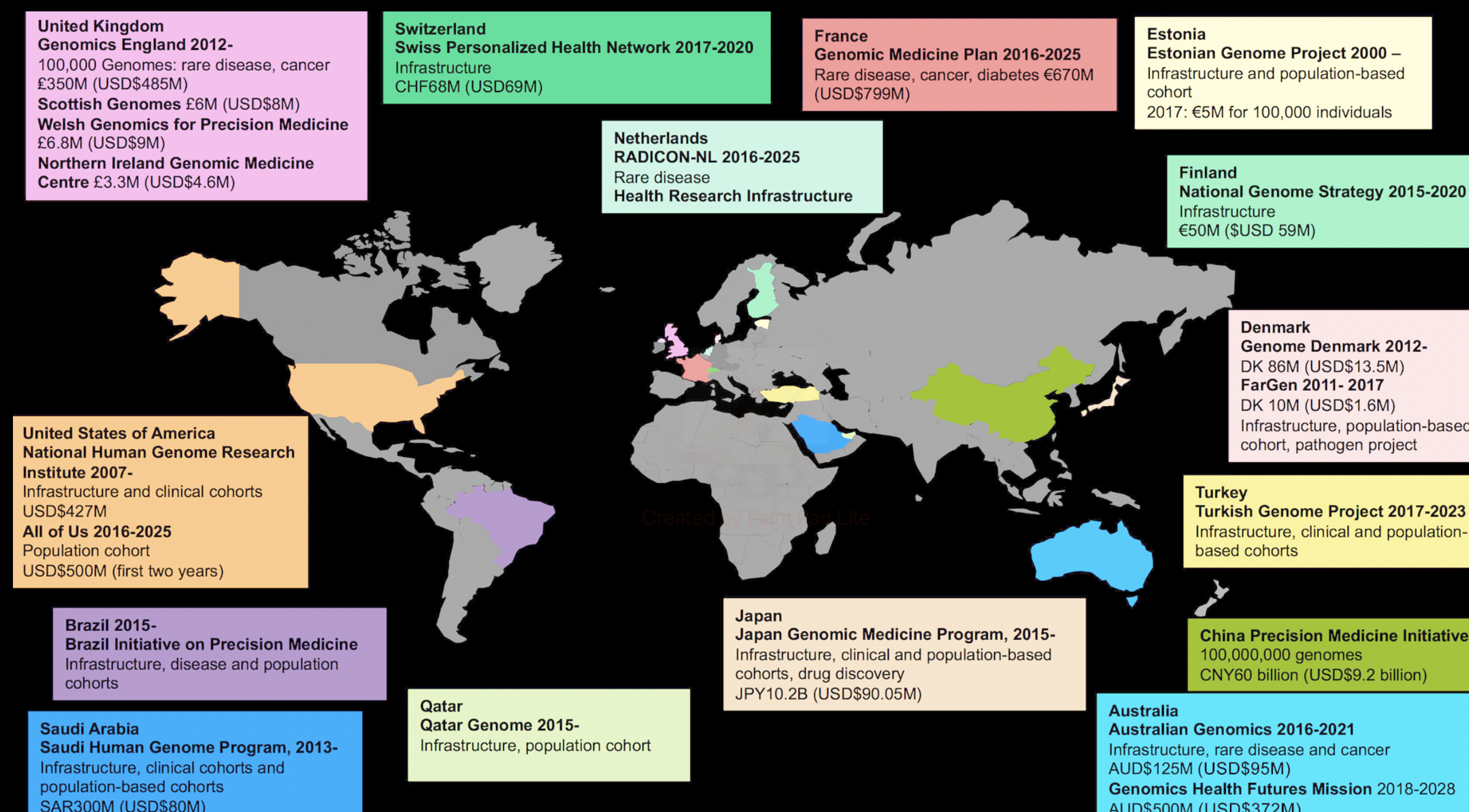
RESULT: REDUCED COMPATIBILITY, PORTABILITY, INTEROPERABILITY, AND INTEGRATION BETWEEN THE TOOLS.

FINDINGS FROM BANI BAKER ET AL. 2020 (DOI.ORG/10.1016/J.IMU.2020.100296)

SECURITY IN DATA COLLABORATIONS

"RESEARCHERS MUST BE ABLE TO RESPONSIBLY INTERROGATE THE WORLD'S COLLECTIVE GENOMIC DATA AS A SINGLE VIRTUAL COHORT THAT TRANSCENDS THE CAPACITY — AND JURISDICTION — OF ANY SINGLE ORGANISATION OR COUNTRY."

FROM ARTICLE - "ELIXIR AND GA4GH EXPAND COLLABORATION", MAY 2019 [HTTPS://WWW.GA4GH.ORG/NEWS/ELIXIR-AND-GA4GH-EXPAND-COLLABORATION](https://www.ga4gh.org/news/elixir-and-ga4gh-expand-collaboration)



STARK ET AL. INTEGRATING GENOMICS INTO HEALTHCARE: A GLOBAL RESPONSIBILITY. THE AMERICAN JOURNAL OF HUMAN GENETICS 104.1 (2019): 13-20.

NEED FOR CLOUD SOLUTIONS

PERSONAL GENOMES CAN BE COMPRESSED AND CONVERTED TO SEARCHABLE FORMATS

CLOUD SIMPLIFIES MAKING DATA F.A.I.R. = **FINDABLE, ACCESSIBLE, INTEROPERABLE AND REUSABLE**

- CONVERT GENOMIC VCF INTO COMPRESSED BIGQUERY OR ATHENA FORMATS
- USE REST API AND SQL TO SEARCH VARIANTS ACROSS GENOMES WITH SIMPLE COORDINATES OR PATTERN MATCHING
- MULTIPLE VARIANT SETS FOR PANGENOMES CAN BE JOINED WITH SIMPLE SQL COMMANDS
- PERSONAL GENOME DATA CAN BE LINK TO PUBLIC TEST DATA SETS SUCH AS 1000 GENOMES, SIMON GENOMES, PLATINUM GENOMES.
- RAW SEQUENCING DATA CAN BE MOVED TO ARCHIVAL STORAGE ONCE ALIGNMENTS ARE PRODUCED.

PERSONAL GENOMES BENEFIT FROM SCALABLE OPTIONS FOR MACHINE LEARNING TO OPTIMIZE COSTS AND REDUCE TIME

KUBEFLOW

- UTILIZES GPU'S (GRAPHICAL PROCESSING UNITS) TO SPEED UP PROCESSING
- PARTITION MULTIPLE JOBS ON ONE GPU
- CAN BE CONTROLLED FROM LOCAL MACHINE OR FROM A VM IN GCP

TENSORFLOW

- ALLOWS EXPANSION OF ML TO TPU'S (TENSOR PROCESSING UNITS)
- HIGHER MEMORY OPTIONS NOW AVAILABLE
- TAKES MINUTES TO COMPLETE VERSUS DAYS OR WEEKS.

PERSONAL GENOMES IN THE CLOUD CAN HAVE AUTOMATED ENCRYPTION AND SECURITY POLICIES

- DEVELOP AND AUTOMATE COMPLIANCE & CYBER RISK MANAGEMENT PROCESSES
- COMPLY WITH FISIM, NIST RMF AND HIPAA
- SECURE GENOMICS DATA, PHI AND CLINICAL\TRANSLATIONAL RESEARCH DATA IN CLOUD ENVIRONMENTS
- AUTOMATE GA4GH, SSO AND CLOUD USAGE POLICIES

onix CONTACT US AT:
HLS@ONIXNET.COM
BIT.LY/ONIXINFO

ONIX ENABLES HEALTH AND LIFE SCIENCE ORGANIZATIONS TO REALIZE THE POWER OF THE CLOUD TO SECURELY WORK SIMPLER, SMARTER AND FASTER IN A HIPAA-COMPLIANT ENVIRONMENT.